2016

Symptom endorsement in men versus women with a diagnosis of depression: A differential item functioning approach

Anna Cavanagh
University of Wollongong, ac316@uowmail.edu.au

Coralie J. Wilson
University of Wollongong, cwilson@uow.edu.au

Peter Caputi
University of Wollongong, pcaputi@uow.edu.au

David J. Kavanagh
Queensland University of Technology, david.kavanagh@qut.edu.au

Symptom endorsement in men versus women with a diagnosis of depression: A differential item functioning approach

Abstract

Background: There is some evidence that, in contrast to depressed women, depressed men tend to report alternative symptoms that are not listed as standard diagnostic criteria. This may possibly lead to an under- or misdiagnosis of depression in men. Aims: This study aims to clarify whether depressed men and women report different symptoms. Methods: This study used data from the 2007 Australian National Survey of Mental Health and Wellbeing that was collected using the World Health Organization's Composite International Diagnostic Interview. Participants with a diagnosis of a depressive disorder with 12-month symptoms (n = 663) were identified and included in this study. Differential item functioning (DIF) was used to test whether depressed men and women endorse different features associated with their condition. Results: Gender-related DIF was present for three symptoms associated with depression. Depressed women were more likely to report 'appetite/weight disturbance', whereas depressed men were more likely to report 'alcohol misuse' and 'substance misuse'. Conclusion: While the results may reflect a greater risk of co-occurring alcohol and substance misuse in men, inclusion of these features in assessments may improve the detection of depression in men, especially if standard depressive symptoms are under-reported.

Keywords

symptom, versus, endorsement, women, diagnosis, depression, differential, item, functioning, approach, men

Disciplines

Education | Social and Behavioral Sciences

Publication Details

Cavanagh, A., Wilson, C. J., Caputi, P. & Kavanagh, D. J. (2016). Symptom endorsement in men versus women with a diagnosis of depression: A differential item functioning approach. International Journal of Social Psychiatry, 62 (6), 549-559.

Title:

Symptom endorsement in men versus women with a diagnosis of depression: A Differential Item Functioning approach

Authors, affiliations:

[1]Anna Cavanagh (corresponding author)

Email: ac316@uowmail.edu.au, Ph: +61 2 4221 4164, Fax: +61 4221 4341

Educational Qualification: Clinical Psychology PhD Candidate

Affiliations: School of Psychology, University of Wollongong

Address: School of Psychology, University of Wollongong, Northfields Avenue, Wollongong, NSW 2522, Australia

[2]Coralie J. Wilson

Email: cwilson@uow.edu.au, Ph: +61 2 4221 5135

Educational Qualification: PhD

Affiliations: Illawarra Health & Medical Research Institute, Graduate School of Medicine, University of Wollongong, Wollongong, NSW, Australia

[3]Peter Caputi

Email: pcaputi@uow.edu.au, Ph: +61 2 4221 3717

Educational Qualification: PhD

Affiliations: Centre for Health Initiatives, School of Psychology, University of Wollongong, Wollongong, NSW, Australia

[4]David J. Kavanagh

Email: david.kavanagh@qut.edu.au, Ph: +61 7 3138 6143

Educational Qualification: PhD

Affiliations: Institute of Health & Biomedical Innovation and School of Psychology & Counselling, Queensland University of Technology, Brisbane, QLD, Australia

Introduction

Previous research has consistently reported that depression is more common in women than men (Kessler et al., 2005; Piccinelli and Wilkinson, 2000). The latest version of the Diagnostic and Statistical Manual of Mental Disorders (DSM 5; American Psychiatric Association, 2013) reports an approximate 2:1 female to male gender ratio of depression prevalence beginning from adolescence. However, there is a growing number of studies that suggests depressed men report alternative symptoms that may not be listed as standard diagnostic criteria for depression, leading to a possible under-diagnosis of male depression cases (Addis, 2008; Emslie et al., 2006, 2007; Martin, Neighbors, & Griffith, 2013). These studies suggest that depressed men are more likely to report externalizing symptoms such as irritability, aggression, and alcohol and substance misuse, known as the 'male depressive syndrome' (Rihmer, Pestality, Pihlgren, & Rutz, 1998), whereas depressed women tend to report internalizing symptoms that are consistent with typical presentations of depression.

In a recent analysis of a nationally representative US sample, depressed men reported higher rates of anger attacks/aggression, substance use and risk taking than depressed women (Martin et al., 2013). In addition, the study found that when these additional, alternative symptoms were included in the assessment of depression, men and women met criteria for depression in equal proportions (Martin et al., 2013). These results suggest that some depressed men may be missed by current assessment instruments, and that depression measures which use only standard diagnostic criteria may be biased towards detecting depression in women, rather than men. Consistent with this view, commonly used self-reported depression measures, such as the Beck Depression Inventory and Center for Epidemiologic Studies Depression Scale, include items that are biased towards including depressed women and excluding depressed men (Salokangas, Vaahtera, Pacriev, Sohlman, & Lehtinen, 2002; Stommel et al., 1993). However, findings remain inconclusive. There is

evidence that alternative symptoms reported to be associated with depressed men are elevated in depressed women too (Möller-Leimkühler, Bottlender, Strauss, & Rutz, 2004), and it has been acknowledged that further research is required to clarify gender-specific symptomatology of depression (Department of Health & Aging, 2010; M. G. Harris et al., 2015; The Royal Australian and New Zealand College of Psychiatrists, 2009).

The item response theory (IRT) method of differential item functioning (DIF) is a way of assessing whether reported gender differences in depression reflect true differences between men and women or may be due to item bias instead. DIF is present when individuals who have the same underlying level of depression show different probabilities of symptom endorsement (Thissen, Steinberg, & Wainer, 1993). DIF has become an increasingly popular method in educational and clinical psychology (Michonski, Sharp, Steinberg, & Zanarini, 2012; Teresi, 2006) and has previously been used to assess differences in depression profiles by gender and by ethnicity (Bares et al., 2012; Carragher et al., 2011; Emmert-Aronson et al., 2014; Kalibatseva et al., 2014; Lange et al., 2002; Uebelacker et al., 2009). For instance, one study assessed gender-based DIF in DSM diagnostic criteria for depression in a large American epidemiological sample and found that depressed women had a significantly higher probability of endorsing the symptom of appetite and weight disturbance than depressed men (Uebelacker et al. 2009). In a similar study that used DIF to explore gender differences in depression in a large Australian sample, men were significantly more likely to endorse the symptom of psychomotor difficulties than women (Carragher et al., 2011). These two studies suggest that there may be somatic differences between men and women that influence their likelihood of endorsing symptoms. Both studies, however, included subclinical cases only and did not assess for gender differences in alternative symptoms of depression.

Although there are existing studies that use the DIF approach, more research on symptom-level based gender differences in depression is warranted. As yet, DIF has not been

used to explore symptom-level gender differences in diagnostic, as well as alternative criteria in participants with a diagnosis of depression. Consequently, the aim of the current study is to evaluate gender-based DIF in a depressed subset of a national community sample, examining standard depression criteria, as well as alternative symptoms that are associated with depression.

Method

Sample and instrument

Data from the 2007 National Survey of Mental Health and Wellbeing (NSMHWB), conducted by the Australian Bureau of Statistics (ABS), were used for the statistical analysis for this paper. The NSMHWB collected information on common mental disorders in the Australian adult population using a stratified, multistage area sampling method to randomly select households from Australian states and territories. Further details about the procedure of the NSMHWB can be found elsewhere (Slade, Johnston, Oakley Browne, Andrews, & Whiteford, 2009). The NSMHWB used the World Health Organisation's Composite International Diagnostic Interview (WMH-CIDI; Kessler & Üstün, 2004) to assess participants. The WMH-CIDI is a fully structured interview that produces diagnoses related to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV/5; American Psychiatric Association, 2013; 2000) and to the International Classification for Diseases (ICD-10; World Health Organisation, 1994). It uses several initial screening questions for the assessment of mental disorders. For depression it includes screening questions relating to core depressive symptoms. If participants answer 'yes' to any of the depressive screening questions, they are further assessed for depressive disorders. All interviewers underwent a comprehensive training program before conducting NSMHWB interviews.

DSM-IV/5 and alternative criteria

DSM-IV and DSM 5 (American Psychiatric Association, 2000, 2013) were used to identify standard criteria for depression. Alternative criteria were identified on the basis of previous evidence in the research literature (Martin et al., 2013; Rice et al., 2015). See Table 1 for a list of alternative criteria included in the study. Items that related to the same criterion were combined into one variable for the main analysis of current study. For instance, items 'wanted to be alone' and 'less talkative' were combined into 'social withdrawal'. Ordinal variables were recoded into categorical variables for primary analyses. Symptom severity for most symptoms could not be obtained as they were measured as either present or absent in the original dataset.

Data analysis strategy

First, chi-square analyses were conducted to evaluate differences in men and women's frequencies of standard and alternative criteria for depression using IBM SPSS Statistics 22 (IBM Corp, 2013). Effect sizes of phi ϕ coefficients for gender differences were interpreted using the following guidelines: 0.0 to 0.1 as small, 0.2 to 0.3 as medium and equal to or above 0.5 as large (Cohen, 1988; Durlak, 2009). Second, DIF was conducted to assess whether men and women differed in their likelihood of endorsing criteria, using the computer software Mplus (Muthén & Muthén, 2009) and IRTPRO (Cai, du Toit, & Thissen, 2011). Both parametric and non-parametric procedures are available for testing DIF and are found to produce generally similar results (Basokcu & Ogretmen, 2014). For this paper, the method of a parametric procedure was used as it allows for an estimation of differences in a fixed set of parameters between groups, which is particularly useful in examining diagnostic statistics (Freedman, 2009). A two-parameter model (2PL), rather than a three-parameter model, was

adopted for the specificity of the 2PL model to estimate parameters in real data sets (D. Harris, 1989). Parameter distribution was checked for each estimated item.
Goodness of fit and unidimensionality were assessed using M_2 statistics and its associated root mean square error approximation (RMSEA) value, as well as standardised local dependence (LD) chi-square indices (Cai et al., 2011; Chen & Thissen, 1997).

LD indices are standardised chi-square values; values above 10.0 challenge the assumption of independency among items (SSI, 2011). For males, none of the standardised chi-square indices of LD was above the recommended value of 10.0. The largest LD was between 'reduced self-esteem/self-confidence' and 'worthlessness/guilt' ($LD\ \chi^2 = 7.1$). For females, there was one LD over the recommended value of 10.0, between 'mad/angry' and 'irritability' ($LD\ \chi^2 = 12.2$). However, as the large majority of standardised chi-square indices of LD were below 10.0, unidimensionality and independence of item pairs could still be assumed for the IRT-DIF analyses.

The 2PL model was fitted to the 19 standard and alternative criteria of depression. Each criterion was dichotomous and rated as either present or absent. For the purpose of this paper, the 2PL model represents the probability of a criterion to be rated as present as a function of the underlying construct of depression liability. For each criterion, discrimination and threshold parameters were estimated. The discrimination parameter, *a*, indicates how well the item differentiates between depressed men and depressed women. The higher the discrimination parameter, the stronger the item's ability to discriminate between genders (DeMars, 2010). The threshold parameter, *b*, refers to item difficulty. Item difficulty reflects the level of depression liability required for men and women to endorse an item with 50% probability. In this study, DIF was considered present when men and women had an unequal probability of endorsing the item that was tested.

DIF analyses were conducted in three steps. In the first step, each item was initially tested for DIF, whilst all other items were used as tentative anchor items. Anchor items comprised items that were believed not to show DIF (Edelen et al., 2006). In the second step, items that showed DIF in the first step were entered into IRTPRO as 'test candidates' and items that did not show DIF were entered as 'anchor' items. In the last step, DIF was examined for each of these test candidates, using the established anchor items. The discrimination parameter (*a*) of these test candidates was constrained to be equal and the threshold parameter (*b*) was assessed separately for men and women. Wald tests, which provide separate statistics for the discrimination and threshold parameters, were used to evaluate the presence of DIF. For all analyses, men were the reference group and women were the focal group.

Results

Sample characteristics

A total of 8841 households took part in the survey, representing a response rate of 60%. Of the 8841 interviewed participants, 663 participants (male $n = 243$; female $n = 420$) were diagnosed with a Minor Depressive Disorder (MND), Major Depressive Disorder (MDD) or Dysthymia (DYS) with 12-month symptoms, and were included in this study. For participant characteristics see Table 1.

Table 1

Participants' socio-demographic characteristics

	Total sample		**Any DSM-IV/5 Depressive Disorder with 12-month symptoms[a]**	
Variable	**Men** **n (%)**	**Women** **n (%)**	**Men** **n (%)**	**Women** **n (%)**
Total	4027 (46%)	4814 (54%)	243 (37%)	420 (63%)
Age (years)				
16-24	681 (17%)	790 (17%)	34 (14%)	83 (20%)
25-34	516 (13%)	774 (16%)	43 (18%)	82 (19%)
35-44	756 (19%)	882 (18%)	62 (25%)	89 (21%)
45-54	566 (14%)	698 (14%)	53 (22%)	75 (18%)
55-64	604 (15%)	669 (14%)	32 (13%)	61 (15%)
65-74	574 (14%)	530 (11%)	14 (6%)	18 (4%)
75-85	319 (8%)	445 (9%)	5 (2%)	11 (3%)
Marital status				

Married/De facto	1935 (48%)	2067 (43%)	61 (25%)	111 (26%)
Widowed/Separated/Divorced	662 (17%)	1283 (27%)	76 (31%)	133 (32%)
Never married	1430 (35%0	1464 (30%)	106 (44%)	176 (42%)
Labour force status				
Employed	2698 (67%)	2801 (58%)	160 (66%)	243 (58%)
Unemployed	98 (2%)	118 (2%)	15 (6%)	20 (5%)
Not in the labour force	1231 (31%)	1895 (39%)	68 (28%)	157 (37%)
Education				
Post-school qualification	2281 (56%)	2492 (51%)	150 (62%)	226 (54%)
No post-school qualification	1676 (42%)	2241 (47%)	91 (37%)	189 (45%)
Unknown	70 (2%)	81 (2%)	2 (1%)	5 (1%)
Country of birth				
Australia	2960 (74%)	3570 (74%)	184 (76%)	327 (78%)
Other English-speaking country	497 (12%)	535 (11%)	27 (11%)	45 (11%)
Other non English-speaking country	570 (14%)	709 (15%)	32 (13%)	48 (11%)

Note.

[a] Includes cases with DSM Minor Depressive Disorder, DSM Major Depressive Disorder or DSM Dysthymia with 12-month symptoms.

Frequency differences by gender

Women were significantly more likely to have a diagnosis of a MND, MDD or DYS with 12-month symptoms than men ($\chi^2(1) = 22.88$, $p < .001$). After controlling for age ($\chi^2(1) = 22.90$, $p < .001$), education ($\chi^2(1) = 22.88$, $p < .001$) and marital status ($\chi^2(1) = 22.90$, $p < .001$), results did not change and women were still significantly more likely to have a diagnosis of depression. Women had significantly higher frequencies on most standard criteria, including appetite/weight disturbance, sleep disturbance, psychomotor disturbance, worthlessness/guilt and suicidality, whereas men had significantly higher frequencies of the alternative criteria of alcohol and substance use (see Table 2). The significant effects were in the small range, except for the alcohol use item ‘drinking caused problems’, which was higher in men and in the medium range, and the depressed mood item ‘often in tears’ which was higher in women and in the medium range.

Table 2

Gender differences in cases with a DSM depression diagnosis with 12-month symptoms[a] (males n=243; females n=420)

	Males	**Females**	**χ^2**	**ϕ[b]**	**ϕ Label[c]**
DSM-IV/5 Criteria for Depressive Disorder					
Depressed mood	96%	98%	0.97	.04	
Feeling so sad could not be cheered up	65%	74%	**6.96****	**.10****	**Small**
Feeling depressed nearly every day	95%	95%	0.00	-.00	
Often in tears	47%	86%	**117.00*****	**.42*****	**Medium**
Diminished interest/pleasure	91%	93%	1.26	.04	
Loss of interest in most activities	84%	88%	2.72	.06	
Loss of pleasure in good things	75%	76%	0.08	.01	
Appetite/weight disturbance	81%	92%	**17.21*****	**.16*****	**Small**
Smaller appetite	63%	62%	0.70	-.01	
Larger appetite	12%	21%	**10.32****	**.13****	**Small**
Weight gain	12%	23%	**12.92*****	**.14*****	**Small**
Weight loss without trying	50%	48%	0.29	-.02	
Sleep disturbance	92%	96%	**4.43***	**.08***	**Small**
Trouble sleeping most nights	82%	86%	1.80	.05	
Sleep a lot more than usual	10%	10%	0.00	.00	
Fatigue/loss of energy					
Feeling tired/low energy	90%	91%	0.06	.01	
Psychomotor disturbance	66%	70%	1.48	.05	
Talk/move more slowly	49%	59%	**5.54***	**.09***	**Small**
Restless	16%	11%	3.22	-.07	
Cognitive difficulties	91%	93%	1.70	.05	

Trouble concentrating	81%	86%	3.37	.07	
Slow thoughts	63%	70%	3.00	.07	
Unable to make up mind	72%	76%	1.31	.04	
Worthlessness/guilt	45%	60%	**13.04*** **	**.14*** **	**Small**
Feel worthless	37%	56%	**20.95*** **	**.18*** **	**Small**
Feel more guilty than should	11%	7%	2.52	-.06	
Reduced self-confidence/self-esteem	87%	89%	1.22	.04	
Loss of self-confidence	82%	87%	2.31	.06	
Feel not as good as other people	62%	70%	**4.59***	**.08***	**Small**
Suicidality	63%	71%	**4.35***	**.08***	**Small**
Thought about death	60%	68%	**4.80***	**.09***	**Small**
Thought about committing suicide	41%	43%	0.24	.02	
Alternative Criteria					
Irritability					
Irritable, grouchy, or in a bad mood	67%	68%	0.10	.00	
Violent urges	40%	37%	0.53	-.03	
Urge to hit, push or hurt someone	27%	24%	0.60	-.03	
Urge to break or smash something	26%	27%	0.85	.01	
Anger/madness					
Feel mad/angry	79%	84%	2.52	.06	
Overwhelmed/stress					
Feel everything was an effort	78%	81%	0.73	.03	
Social withdrawal	90%	93%	2.00	.05	
Wanted to be alone	78%	81%	0.61	.30	
Less talkative	84%	86%	0.22	.02	
Functional impairment	36%	46%	**6.50***	**.10***	**Small**
Interference of episode with work/social/personal relations	6%	8%	0.22	.30	

Inability to do daily activities in worst episode	75%	81%	1.73	.07	
Anxiety	79%	78%	0.02	-.01	
Nervous or anxious	5%	4%	0.39	-.02	
Worry more than other people	73%	74%	0.03	.01	
Alcohol use	35%	19%	**21.91*** **	**-.18*** **	**Small**
Drank more frequently than intended	9%	6%	1.12	-.04	
Drinking caused problems with family/friends/others	31%	14%	**26.10*** **	**-.20*** **	**Small**
Substance use[d]	20%	12%	**8.83****	**-.12****	**Small**
Used drugs more frequently than intended	7%	5%	1.99	-.06	
Drug use caused problems with family/friends/others	17%	8%	**13.45*** **	**-.14*** **	**Small**
Used drugs to keep from having problems	6%	3%	**4.33***	**-.08***	**Small**

Note.

[a] This includes cases with a diagnosis of a DSM Minor Depressive Disorder, DSM Major Depressive Disorder or DSM Dysthymia with 12-month symptoms.

[b] Indicates effect size Phi ϕ.

[c] Small effect refers to Phi ϕ = 0.1, medium effect refers Phi ϕ = 0.3, large effect refers to Phi ϕ = 0.5 (Cohen, 1988).

[d] Substances include marijuana, stimulants, sedatives or opioids.

* $p < .05$

** $p < .01$

*** $p < .001$

Item Response Models

Unidimensionality

Model fit was analysed separately for males and females. For males, the 2PL unidimensional IRT model showed satisfactory fit: M_2 (152) = 266.78, $p < .001$; RMSEA < 0.06. For females, the 2PL unidimensional IRT model also showed satisfactory fit: M_2 (152) = 362.01, $p < .001$; RMSEA < 0.06. Although the M_2 statistic was significant indicating some model error, the RMSEA values below 0.06 indicate adequate model fit (Hu & Bentler, 1995).

Detection of DIF

The first step in the DIF analyses was to identify anchor and candidate items. Each item was initially tested for DIF, while using the other items as anchors. Fifteen items were identified as anchors that showed no DIF indicated by non-significant Wald test (χ^2) statistics ($p > .05$). The remaining items 'appetite/weight disturbance' ($p = .0021$), 'worthlessness/guilt' ($p = .0302$), 'alcohol use' ($p = .0001$), and 'substance use' ($p = .0006$) were significant. Items were further evaluated for DIF in a second separate analysis to confirm whether they function appropriately as DIF and anchor items. 'Worthlessness/guilt' ($p = 0.1147$) did not exhibit DIF and was entered as an anchor item instead. The remaining three test items exhibited DIF. For each item, DIF was concentrated in the threshold (b) parameter as indicated by significant Wald tests: For 'appetite/weight disturbance' χ^2 (1) = 8.2, $p < .01$; for 'alcohol use' (χ^2 (1) = 27.9, $p < .001$) and for 'substance use' (χ^2 (1) = 13.5, $p < .001$). In the final analysis, the discrimination and threshold parameters for the anchor items were constrained to be equal for both genders. For the DIF items, only the discrimination parameters were constrained to be equal across gender so that threshold parameters could be estimated freely, see Table 3.

Table 2

IRT Item Parameter Estimates

Items	**Gender**	***a*[1] *(standard error)***	***b*[2] *(standard error)***
Anchor items			
Depressed mood	Both	1.02 (0.33)	-3.69 (0.99)
Diminished interest / pleasure	Both	1.47 (0.28)	-2.03 (0.28)
Sleeping disturbance	Both	0.88 (0.23)	-3.33 (0.78)
Psychomotor disturbance	Both	0.90 (0.16)	-0.77 (0.17)
Fatigue / loss of energy	Both	0.49 (0.18)	-4.53 (1.57)
Worthlessness / guilt	Both	1.36 (0.21)	0.05 (0.10)
Reduced self-confidence / self-esteem	Both	1.39 (0.24)	-1.68 (0.23)
Suicidality	Both	0.88 (0.16)	-0.78 (0.17)
Cognitive difficulties	Both	1.28 (0.25)	-2.21 (0.34)
Anger / madness	Both	0.34 (0.15)	-4.42 (1.89)
Violent urges	Both	0.47 (0.13)	1.36 (0.37)
Social withdrawal	Both	1.11 (0.23)	-2.37 (0.40)
Functional impairment	Both	0.75 (0.14)	0.71 (0.17)
Irritability	Both	0.62 (0.14)	-1.07 (0.28)
Overwhelmed / stress	Both	0.87 (0.17)	-1.60 (0.30)
Anxiety	Both	0.55 (0.14)	-2.26 (0.59)
Threshold DIF items[3]			
Appetite / weight disturbance	Males	1.13 (0.27)	-1.26 (0.30)
	Females	1.13 (0.27)	-1.80 (0.50)
Alcohol use	Males	1.18 (0.24)	0.83 (0.18)
	Females	1.18 (0.24)	1.84 (0.30
Substance use	Males	1.46 (0.26)	1.43 (0.23)
	Females	1.46 (0.26)	2.07 (0.29)

Note.

[1] Refers to discrimination parameter *a*.
[2] Refers to threshold parameter *b*.
[3] For threshold DIF items, the estimates of the discrimination parameter *a* were constrained equal for men and women and the estimates for threshold parameter *b* were assessed separately for men and women.

DIF items

Three items showed significant DIF in the threshold parameters across gender, see Table 3. Women had a significantly higher probability of endorsing 'appetite/weight disturbance' than men. Specifically, men were 0.54 standard units higher in the level depression liability required to have a 50% chance of endorsing this item. Conversely, men had a significantly higher probability of endorsing the items 'alcohol use' and 'substance use' than women. This means women were 1.01 standard units higher in the level of depression liability required to have a 50% chance of presenting with 'alcohol use' and 0.64 standard units higher in the level of depression liability required to have a 50% chance of presenting with 'substance use'.

Anchor items

Anchor items that had discrimination parameters of value 1 or greater were considered to be adequately discriminating (Holland & Wainer, 1993). These items included 'depressed mood', 'diminished interest interest/pleasure', 'worthlessness/guilt', 'reduced self-confidence', and 'social withdrawal', see Table 3. The direction of most threshold parameters indicates that women were more likely to endorse items than men, except for 'worthlessness/guilt', 'violent urges' and 'functional impairment', where the reverse was true.

Test curves

The test characteristic curves provide information on the impact of DIF on the performance of the overall set of items (Figure 1). The curves represent the expected summed scores (i.e. the expected sum of all of the 19 standard and alternative criteria) as a function of participants' liability to have a diagnosis of depression. Figure 1 shows that men are expected

to be rated slightly higher than women on the WMH-CIDI at the same level of depression liability. Figure 2 depicts the test information curves for both genders for the test as a whole. These curves provide information about the preciseness of measuring the underlying construct of depression liability. As displayed in Figure 2, measurement of depression is most precise starting at the negative end and ranging to the positive end of the continuum, within the range of -2.5 to +2.0.

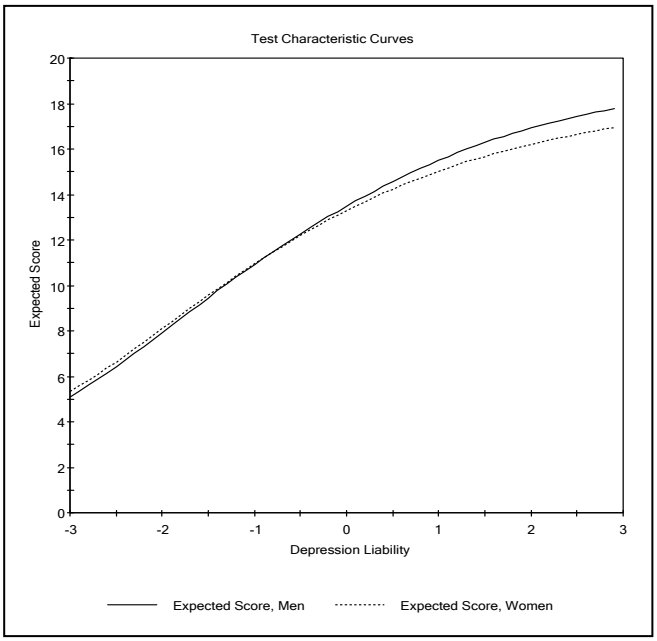

Figure 1. Test characteristic curves by gender

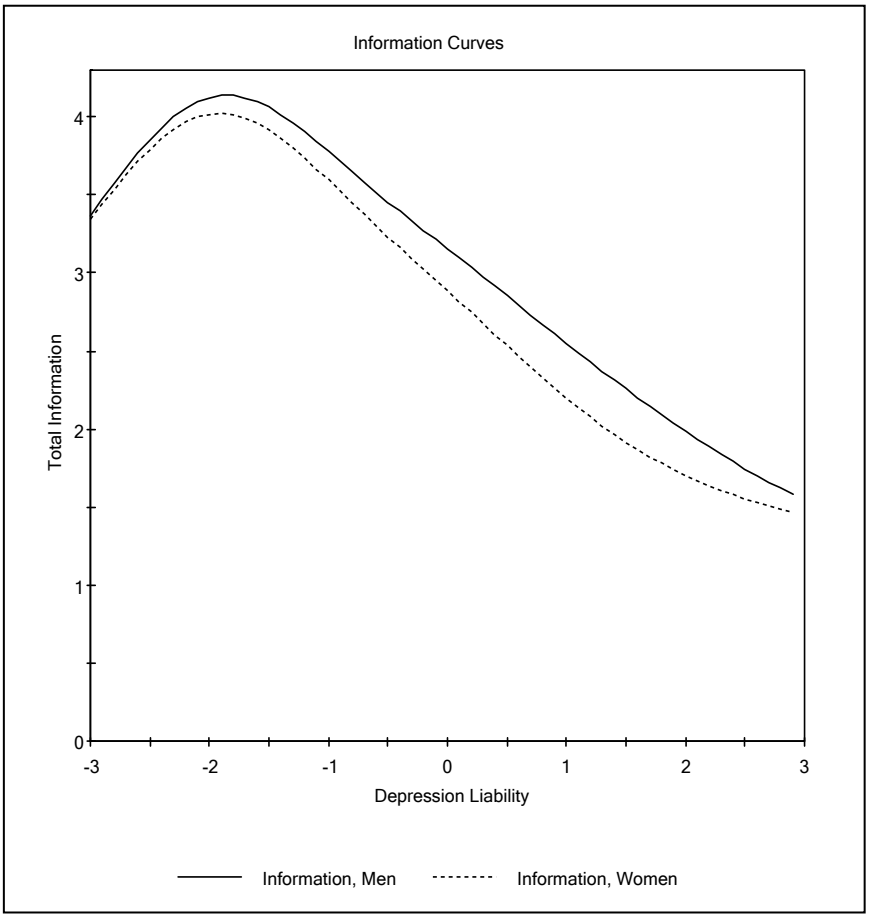

Figure 2. Information curves by gender

Discussion

The current study evaluated gender differences in standard and alternative symptoms, in participants with a DSM diagnosis of a depressive disorder recruited from a large, national community sample. Consistent with previous research on depression prevalence, this study found that significantly higher frequencies of women had a DSM diagnosis of a depressive disorder than men. Furthermore, preliminary findings indicate that women had higher frequencies of symptoms that are part of DSM criteria for depression than men. However, the size of these differences was small except for the depressed mood item ‘often in tears’, where the effect size was in the medium range. Additionally, women had slightly higher frequencies

of alternative symptoms associated with depression, except for the symptoms of violent urges and alcohol and substance use, which were more frequent in men.

DIF analyses showed that except for three symptoms, most standard and alternative symptoms functioned similarly for men and women with a DSM diagnosis of a depressive disorder. Depressed women had a higher likelihood of endorsing the symptom appetite/weight disturbance, whereas depressed men had a higher likelihood of endorsing the symptoms of alcohol and substance use. Particularly, depressed women were significantly more likely to report an increase in appetite and weight, in comparison to men who had slightly higher frequencies of reporting a decrease in appetite and weight loss without trying. This finding of higher appetite/weight disturbance in women is consistent with a previous study that used DIF in a sub-clinical sample (Uebelacker et al., 2009), as well as with a study that assessed gender differences in a clinical subset of a general population sample (Silverstein, 2002). The reasons underlying this finding are unclear, but it may be that biological and hormonal differences between men and women influence their eating behaviours and metabolic processes. Another possibility may be that gender differences in appetite/weight disturbance relate to gender differences in social expectations around body image and appearance, which is also reflected in the higher frequencies of women experiencing eating disorders, in comparison to men (Silverstein, 2002; Uebelacker et al., 2009).

Although the finding that appetite/weight disturbance was more frequent in depressed women than depressed men was consistent with some studies in the research literature, it diverged from a similar study that also used DIF to assess gender differences in DSM criteria for depression in an Australian national sample (Carragher et al., 2011). Carragher et al. (2011) did not find gender-related DIF for appetite/weight disturbance but for psychomotor difficulties. They found that men were more likely to endorse this symptom than women. One

possible explanation for these differences in findings may be that Carragher et al. (2011) included sub-clinical cases of men and women, whereas the current study included participants with a diagnosis of a depressive disorder. It may be that gender differences in symptoms differ between sub-clinical and clinical depression cases. However, Carragher et al. (2011) and the current study are consistent in that they both found that gender differences were small. Nevertheless, it may be important for clinicians to be mindful of possible gender differences in symptoms related to somatic processes when assessing depression. Specifically, an increase in appetite and weight may be a more sensitive risk indicator for mild to moderately severe depression in women, whereas a decrease in appetite and weight may be a more sensitive risk indicator for lower levels of depression severity in men.

The finding that alcohol and substance misuse was more frequent in depressed men than depressed women is coherent with studies indicating higher alcohol and substance use in men with depression (Martin et al., 2013; Satre et al., 2011). Depressed men's higher likelihood of reporting alcohol and substance misuse may be related to men's general higher incidence rates of alcohol and substance misuse in comparison to women (Brady and Randall, 1999; Kessler et al., 1994). Alternatively, it is possible that alcohol and substance misuse is more intimately linked to depression in men, as a way of self-medication or coping, than in women (Whittle et al., 2015). Differential societal pressures and expectations for men and women are also likely to play a role in the perception and expression of symptoms (Danielsson & Johansson, 2005; Emslie et al., 2006; Ridge, Emslie, & White, 2011). For instance, depressed men may feel that an increase in alcohol and other substance use is a more gender-appropriate way of expressing their distress than through sad mood or crying. In fact, alcohol and other substances may effectively substitute for eating as a more gender-appropriate self-indulgent response to depression in men. Further investigation of the role that alcohol and substance misuse play in male depression should be undertaken, especially

since misuse of alcohol and other substances exacerbates depressive symptoms (Bolton, Robinson, & Sareen, 2009).

Unlike other studies (Genuchi and Mitsunaga, 2015; Martin et al., 2013), the current study does not support the hypothesis of a distinctively male sub-type of depression characterised by irritability, aggression and violent tendencies. Rather, these symptoms are also elevated in depressed women. The contention that they may play a role in depression for both genders is further supported by evidence of unidimensionality across the examined criteria, suggesting that they reflect a single underlying construct of depression. Future research is required to further assess the role of alternative depression symptoms in both genders.

While this study does not support the existence of a separate 'male depressive syndrome', findings suggest that there are some differences in the presentation of symptoms in depressed men versus depressed women that may be important for clinicians to consider in their assessments. Assessments that include gender-sensitive questioning or the use of gender-specific depression instruments (e.g. Rice et al., 2013) may lead to an increased number of men and women identified at risk for depression. Beyond clinical utility, findings indicate that it may be particularly important for education and policy to consider gender and specifically target the needs of men and women. Men might particular benefit from interventions that are practical and behavioural-based, whereas women might be more likely to benefit from more traditional health-based interventions.

Limitations

Given that a secondary analysis of an already existing dataset was conducted in this study, not all research variables reported that may be relevant to depression in men were investigated. For instance, depressed men are more likely to engage in risk taking behaviours,

such as gambling or drink driving, than depressed women (Brownhill, Wilhelm, Barclay, & Schmied, 2005; Martin et al., 2013). However, these behaviours could not be assessed as the dataset for the current study did not include items that adequately described or related to the construct of risk taking. Furthermore, some variables included in the analysis of this study were only measured using one or two items. For example, irritability was measured by the single item 'irritable, grouchy or in a bad mood' as part of the depression section. The measurement of irritability by a single item could have potentially led to an under or mis-identification of 'true' irritability, such as caused by the ambiguity of the term 'bad mood' (Kovess-Masfety et al., 2013).

It needs to be taken into consideration that the 2007 dataset used for the paper was collected almost 10 years ago. Considerate effort to better understand male mental health has occurred since then, implying that the results of this study might possibly not reflect these changes. This highlights the urgency of re-estimating findings in a more timely national dataset.

In common with other research on this topic, our results are also potentially subject to under-reporting of symptoms by men, because of social desirability, wording of particular items, lack of emotional awareness or inappropriate emotional regulation strategies (Barrett et al., 2000; Cavanagh et al., in press; Lange et al., 2002; Nolen-Hoeksema, 2011).

Due to the structure of the diagnostic tool used for this study, only participants that endorsed the depression screening questions of depressed mood and/or loss of interest were tested on the remaining standard depression items and included in the current study (Carragher et al., 2011). Depressed men, however, may be less likely to report depressed mood or loss of interest than depressed women (Salokangas et al., 2002; Stommel et al., 1993). The structure of the diagnostic tool used for this study may therefore have led to the exclusion of men who would have endorsed other items and should be considered as being

depressed in a diagnostic system that was more sensitive to the presentation of depression in men. This study estimated endorsement of symptoms in a clinical subset of men and women. It may be of interest for future research to investigate gender differences in symptom endorsement in a sub-clinical dataset.

Conclusions

Gender-related DIF was present for the symptoms of appetite/weight disturbance and alcohol and substance misuse. Women who experience depression may more readily report an increase in appetite and weight, whereas the reverse is true for men. In contrast, depressed men more easily endorse alcohol and substance misuse than depressed women. Considering these gender differences in symptom endorsement may be important for clinicians when assessing and treating depression in order improve the detection of depression in both genders.

Role of the Funding Source

This work was supported by an Australian Postgraduate Award provided by the University of Wollongong to the first author.

Declaration of Conflicting Interests

All authors declare that they have no conflict of interest.

References

Addis, M. E. (2008). Gender and depression in men. *Clinical Psychology: Science and Practice*, *15*(3), 153–168. http://doi.org/10.1111/j.1468-2850.2008.00125.x

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders (4th ed., text rev.)*. Arlington, VA: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Washington, DC: American Psychiatric Press.

Bares, C., Andrade, F., Delva, J., Grogan-Kaylor, A., & Kamata, A. (2012). Differential item functioning due to gender between depression and anxiety items among Chilean adolescents. *International Journal of Social Psychiatry*, *58*(4), 386–392. http://doi.org/10.1177/0020764011400999

Basokcu, T. O., & Ogretmen, T. (2014). Comparison of parametric and nonparametric item response techniques in determining Differential Item Functioning in polytomous scale. *American Journal of Theoretical and Applied Statistics*, *3*(2), 31.

http://doi.org/10.11648/j.ajtas.20140302.11

Bolton, J. M., Robinson, J., & Sareen, J. (2009). Self-medication of mood disorders with alcohol and drugs in the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Affective Disorders*, *115*(3), 367–375. http://doi.org/10.1016/j.jad.2008.10.003

Brady, K. T., & Randall, C. L. (1999). Gender differences in substance use disorders. *The Psychiatric Clinics of North America*, *22*(2), 241–252. http://doi.org/10.1016/S0193-953X(05)70074-5

Brownhill, S., Wilhelm, K., Barclay, L., & Schmied, V. (2005). "Big build": hidden depression in men. *The Australian and New Zealand Journal of Psychiatry*, *39*(10), 921–31. http://doi.org/10.1111/j.1440-1614.2005.01665.x

Cai, L., du Toit, S. H. C., Thissen, D. (2011). IRTPRO: Flexible professional item response theory modeling for patient reported outcomes. [Computer software]. Chicago, IL: SSI International.

Carragher, N., Mewton, L., Slade, T., & Teesson, M. (2011). An item response analysis of the DSM-IV criteria for major depression: Findings from the Australian National Survey of Mental Health and Wellbeing. *Journal of Affective Disorders*, *130*(1-2), 92–8. http://doi.org/10.1016/j.jad.2010.09.033

Cavanagh, A., Caputi, P., Wilson, C., & Kavanagh, D. (in press). Gender differences in self-reported depression and co-occurring anxiety and stress in a vulnerable community population. *Australian Psychologist*.

Chen, W.-H., & Thissen, D. (1997). Local Dependence Indexes for Item Pairs Using Item Response Theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265–289. http://doi.org/10.3102/10769986022003265

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Erlbaum.

Danielsson, U., & Johansson, E. E. (2005). Beyond weeping and crying: a gender analysis of expressions of depression. *Scandinavian Journal of Primary Health Care*, *23*(3), 171–7. http://doi.org/10.1080/02813430510031315

DeMars, C. (2010). *Item Response Theory*. New York, NY: Oxford University Press.

Department of Health & Aging. (2010). *National Male Health Policy: Building on the strengths of Australian males*. Canberra: Commonwealth of Australia.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*(9), 917–928. http://doi.org/10.1093/jpepsy/jsp004

Emmert-Aronson, B. O., Moore, M. T., & Brown, T. A. (2014). Differential Item Functioning of the Symptoms of Major Depression by Race and Ethnicity: An Item Response Theory Analysis. *Journal of Psychopathology and Behavioral Assessment*, *36*(3), 424–431. http://doi.org/10.1007/s10862-014-9407-4

Emslie, C., Ridge, D., Ziebland, S., & Hunt, K. (2006). Men's accounts of depression: Reconstructing or resisting hegemonic masculinity? *Social Science and Medicine*, *62*(9), 2246–2257. http://doi.org/10.1016/j.socscimed.2005.10.017

Emslie, C., Ridge, D., Ziebland, S., & Hunt, K. (2007). Exploring men's and women's experiences of depression and engagement with health professionals: more similarities than differences? A qualitative interview study. *BMC Family Practice*, *8*, 43.

http://doi.org/10.1186/1471-2296-8-43

Freedman, D. A. (2009). *Statistical Models: Theory and Practice* (Second). Cambridge, England: Cambridge University Press.

Genuchi, M. C., & Mitsunaga, L. K. (2015). Sex differences in masculine depression: Externalizing symptoms as a primary feature of depression in men. *The Journal of Men's Studies*, *23*(3), 243–251. http://doi.org/10.1177/1060826514561986

Harris, D. (1989). Comparison of 1-,2- and 3-parameter IRT Models. *Educational Measurement: Issues and Practice*, *8*(1), 35–41. http://doi.org/10.1111/j.1745-3992.1989.tb00313.x

Harris, M. G., Diminic, S., Reavley, N., Baxter, A., Pirkis, J., & Whiteford, H. A. (2015). Males' mental health disadvantage: An estimation of gender-specific changes in service utilisation for mental and substance use disorders in Australia. *Australian & New Zealand Journal of Psychiatry*, *49*(9), 821–832. http://doi.org/10.1177/0004867415577434

Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum.

Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

IBM Corp. (2013). IBM SPSS statistics for windows (Version 22.0). [Computer software]. Armonk, NY: IBMCorp.

Kalibatseva, Z., Leong, F. T. L., & Ham, E. H. (2014). A symptom profile of depression among Asian Americans: is there evidence for differential item functioning of depressive symptoms? *Psychological Medicine*, *44*(12), 2567–78. http://doi.org/10.1017/S0033291714000130

Kessler, T., & Üstün, R. C. (2004). *The WHO World Mental Health Survey: global perspectives on the epidemiology of mental disorders.* Geneva: Cambridge University Press.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*, *62*(June), 593–602. http://doi.org/10.1001/archpsyc.62.6.593

Kovess-Masfety, V., Alonso, J., Angermeyer, M., Bromet, E., de Girolamo, G., de Jonge, P., … Kessler, R. C. (2013). Irritable mood in adult major depressive disorder: Results from the world mental health surveys. *Depression and Anxiety*, *30*(4), 395–406. http://doi.org/10.1002/da.22033

Lange, R., Thalbourne, M. A., Houran, J., & Lester, D. (2002). Depressive response sets due to gender and culture-based Differential Item Functioning. *Personality and Individual Differences*, *33*(6), 937–954. http://doi.org/10.1016/S0191-8869(01)00203-3

Martin, L. A., Neighbors, H. W., & Griffith, D. M. (2013). The experience of symptoms of depression in men vs women: Analysis of the National Comorbidity Survey Replication. *JAMA Psychiatry*, *70*(10), 1100–6. http://doi.org/10.1001/jamapsychiatry.2013.1985

Michonski, J. D., Sharp, C., Steinberg, L., & Zanarini, M. C. (2012). An Item Response Theory Analysis of the DSM–IV Borderline Personality Disorder Criteria in a

Population-Based Sample of 11- to 12-Year-Old Children. *Personality Disorders: Theory, Research, and Treatment*, *4*(1), 15–22. http://doi.org/10.1037/a0027948

Möller-Leimkühler, A. M., Bottlender, R., Strauss, A., & Rutz, W. (2004). Is there evidence for a male depressive syndrome in inpatients with major depression? *Journal of Affective Disorders*, *80*(1), 87–93. http://doi.org/10.1016/S0165-0327(03)00051-X

Muthén, B. O., & Muthén, L. (2009). Mplus (Version 5.21). [Computer software]. Los Angeles, CA: Muthén & Muthén.

Piccinelli, M., & Wilkinson, G. (2000). Gender differences in depression. Critical review. *British Journal of Psychiatry*, *177*(DEC.), 486–492. http://doi.org/10.1192/bjp.177.6.486

Rice, S. M., Fallon, B. J., Aucote, H. M., & Möller-Leimkühler, A. M. (2013). Development and preliminary validation of the male depression risk scale: Furthering the assessment of depression in men. *Journal of Affective Disorders*, *151*, 950–958. http://doi.org/10.1016/j.jad.2013.08.013

Rice, S. M., Fallon, B. J., Aucote, H. M., Möller-Leimkühler, A. M., Treeby, M. S., & P, A. G. (2015). Longitudinal sex differences of externalising and internalising depression symptom trajectories: Implications for assessment of depression in men from an online study. *International Journal of Social Psychiatry*, *61*(3), 236–240. http://doi.org/10.1177/0020764014540149

Ridge, D., Emslie, C., & White, A. (2011). Understanding how men experience, express and cope with mental distress: Where next? *Sociology of Health and Illness*, *33*(1), 145–159. http://doi.org/10.1111/j.1467-9566.2010.01266.x

Rihmer, Z., Pestality, P., Pihlgren, H., & Rutz, W. (1998). "Anxiety/aggression-driven depression" and "male depressive syndrome": Are they the same? *Psychiatric Research*, *77*, 209–210. http://doi.org/10.1080/13518040701205365

Salokangas, R. K. R., Vaahtera, K., Pacriev, S., Sohlman, B., & Lehtinen, V. (2002). Gender differences in depressive symptoms. An artefact caused by measurement instruments? *Journal of Affective Disorders*, *68*(2-3), 215–20. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12063149

Satre, D. D., Chi, F. W., Eisendrath, S., & Weisner, C. (2011). Subdiagnostic alcohol use by depressed men and women seeking outpatient psychiatric services: Consumption patterns and motivation to reduce drinking. *Alcoholism: Clinical and Experimental Research*, *35*(4), 695–702. http://doi.org/10.1111/j.1530-0277.2010.01387.x

Silverstein, B. (2002). Gender differences in the prevalence of somatic versus pure depression: A replication. *American Journal of Psychiatry*, *159*(6), 1051–1052. http://doi.org/10.1176/appi.ajp.159.6.1051

Slade, T., Johnston, A., Oakley Browne, M. A., Andrews, G., & Whiteford, H. (2009). 2007 National Survey of Mental Health and Wellbeing: methods and key findings. *The Australian and New Zealand Journal of Psychiatry*, *43*(7), 594–605. http://doi.org/10.1080/00048670902970882

SSI. (2011). IRTPRO user's guide. Lincolnwood, IL: Scientific Software International, Inc.

Stommel, M., Given, B. A., Given, C. W., Kalaian, H. A., Schulz, R., & Mccorkle, R. (1993). Gender bias in the measurement properties of the Center for Epidemiologic Studies Depression scale (CES-D). *Psychiatry Research*, *49*, 239–250.

Teresi, J. A. (2006). Different approaches to differential item functioning in health applications. Advantages, disadvantages and some neglected topics. *Medical Care*, *44*(11 Suppl 3), S152–S170. http://doi.org/10.2307/41219515

The Royal Australian and New Zealand College of Psychiatrists. (2009). *Submission to the Inquiry of the Senate Select Committee on Men's Health*. Melbourne, VIC. Retrieved from https://www.ranzcp.org/Files/Resources/Submissions/sub14-pdf.aspx

Thissen, D., Steinberg, L., & Wainer, H. (1993). Differential item functioning. In P. W. Holland & H. Wainer (Eds.), (pp. 67–113). Hillsdale, NJ: Lawrence Erlbaum.

Uebelacker, L. A., Strong, D., Weinstock, L. M., & Miller, I. W. (2009). Use of item response theory to understand differential functioning of DSM-IV major depression symptoms by race, ethnicity and gender. *Psychological Medicine*, *39*(4), 591–601. http://doi.org/10.1017/S0033291708003875

Whittle, E. L., Fogarty, A. S., Tugendrajch, S., Player, M. J., Wilhelm, K., Hadzi-pavlovic, D., … Proudfoot, J. (2015). Men, depression, and coping: Are we on the right path? *Psychology of Men & Masculinity*. http://doi.org/http://dx.doi.org/10.1037/a0039024

World Health Organization. (1994). *ICD-10: International Statistical Classification of Diseases and related Health Problems*. Geneva: World Health Organization.